# Optimising ex situ genetic resource collections for European livestock conservation

de Oliveira Silva, R; Vosough Ahmadi, B; Hiemstra, SJ; Moran, D

*Published in:*
Journal of Animal Breeding and Genetics

*DOI:*
10.1111/jbg.12368

First published: 11/12/2018

*Document Version*
Peer reviewed version

**Title: Optimising *ex situ* genetic resource collections for European livestock conservation**

**Running title: Optimising *ex situ* genetic collections**

Rafael De Oliveira Silva[1], Bouda Vosough Ahmadi[2], Sipke Joost Hiemstra[3], Dominic Moran[1]

[1] Global Academy of Agriculture and Food Security, the University of Edinburgh, Easter Bush Campus, Midlothian, EH25 9RG, Edinburgh, UK

[2] Land Economy, Environment and Society Research Group, Scotland's Rural College (SRUC), West Mains Road, EH9 3JG, Edinburgh, UK

[3] Centre for Genetic Resources, Wageningen University & Research, Wageningen Campus, Droevendaalsesteeg 1, 6700 AH Wageningen, The Netherlands

Corresponding author: mathros@gmail.com

## Abstract

*Ex situ* collections offer the potential to reduce extinction risks, affording option to society in maintaining future breeding opportunities for productivity and heritage traits. However, how much should we be seeking to collect and conserve in gene banks, and where? We developed a mathematical model to optimise logistical decisions of breed conservation choices and to evaluate alternative scenarios for efficiently re-allocating genetic materials currently stored in different European gene banks, allowing for cross-country collections, cost and cryogenic capacity differentials. We show how alternative allocations for the breeds that are currently stored in 11 European gene banks could reduce overall conservation costs by around 20% by selecting cryogenic banks that have relatively lower combination of fixed and collection costs, and are geographically closer to collection regions. Our results show that centralizing collections in one gene bank would double the costs, relative to collective European collections approaches. We also calculate marginal costs of collections and show that increasing diversity within the gene banks implies in higher costs per conserved breed.



**Keywords:** *Ex situ* conservation, livestock diversity, optimisation, gene bank, cryoconservation

## Introduction

The increasing vulnerability of *in situ* animal and plant genetic resources for agriculture has been suggested by many authors, emphasising the threats posed by climate change and increasing homogenisation of farming and food production systems (FAO, 2015). The resilience of animal and plant varieties used for agriculture can potentially be maintained by *ex situ* collection of genetic and reproductive materials that can be used to improve and potentially to restore breeds. Gene banks complement *in situ* conservation, and include formal and informal use and exchange of genomic (e.g. DNA, blood, tissue) and reproductive germplasm (e.g. semen, embryos). Globally there are many agricultural biobank collections, typically held for specialised productive purposes. Some national and regional specialised collections emphasise indigenous and cultural breed attributes, e.g. Rare Breeds Survival Trust (http://www.rbst.org.uk) in the UK. Other national and global plant and animal collections are held as public good resources in networks under The International Treaty on Plant Genetic Resources for Food and Agriculture (FAO, 2009), or the centres of Consultative Group on International Agricultural Research (http://www.cgiar.org). But other national and regional collections are more proprietary, offering restricted access usually through subscription. More generally, the academic literature on *ex situ* conservation is skewed towards storage of plant materials. There have been limited efforts to detail or audit animal collections. The European Gene bank Network for Animal Genetic Resources (EUGENA), coordinated by the European Regional Focal Point on Animal Genetic Resources (http://www.rfp-europe.org), is an emerging networking activity specifically targeting national farm animal genetic resource

collections (Hiemstra, Martyniuk, Duchev, & Begemann, 2014). Furthermore, a recent survey conducted as part of the EU IMAGE (http://www.imageh2020.eu/) collaborative project (Passemard et al., 2018) elicited responses from 62 European organizations in 21 countries. The survey revealed some 30 genomic and 51 germplasm collections, with 20% of the organizations holding both germplasm and genomic materials. The data showed over- representation of some countries (e.g. Spain with 26 germplasm and 7 genomic collections).

Beyond Europe, The United States Department of Agriculture supports a National Animal Germplasm Program storing genetic material for use by industry and the research community (Blackburn, 2009). However, collection does not appear to be guided by any clear economic criteria beyond a budget constraint. FAO (2015) collects global data on stored genetic materials of various breeds, but depends on voluntary country reports that are often incomplete. Groeneveld et al. (2016) reviewed bio banking effort for all species and note a lack of a unified and generalised approach to sample collections in the domesticated animal sector.

There is general agreement that *ex situ* collections offer option value, i.e., value of preserving a back-up collection of (threatened) breeds so that this genetic diversity might be available for use in the future. But the efficacy of collections is also largely anecdotal, with some concern that materials stored in gene banks may be compromised, or become redundant or mismatched with independently evolved *in situ* conditions (McGowan et al., 2017). A further caveat on option value is the extent of overlap and possible

redundancy in collections covering countries of similar agri-climatic systems. In a collaborative system of material exchange this redundancy increases the cost of supplying diversity (Blackburn, 2009). While there is considerable focus on the efficiency of *in situ* biodiversity conservation, including area selection algorithms for systematic conservation planning (Kukkala & Moilanen, 2013; Önal, 2003) or conservation funds allocation (Reist-Marti et al., 2006) we are unaware of work optimising *ex situ* livestock collections. Specifically, the *ex situ* literature is apparently limited to optimising genetic variability; i.e. which breeds to conserve. However, as noted by Blackburn (2009) the logistical dimension of collections is an important but neglected limiting factor.

This paper develops a Mixed Integer Programming (MIP), a type of optimization model consisting of both integer (stored breeds) and continuous variables (e.g., costs) as opposed to Linear Programing (LP) which contains only continuous variables (Lee & Letchford, 2007). We use the model to identify the least cost collection and storage strategies for European livestock breeds under a collective budget constraint, and allowing cross-country collections. We use the analysis to construct diversity supply curves to illustrate the relationship between cost and diversity in hypothetical rationalised *ex situ* collections. The analysis goes beyond existing *ex situ* cost exercises that have not considered the efficiency of potentially rationalised collections in a collaborative network (Pizzi et al., 2016). The paper is structured as follows. Section two sets out the theoretical model for rationalisation of collection effort, the survey data used in the

104 analysis, and develops some scenarios for breed collection. Sections three and four
105 provide results and discussion respectively, while section five offers conclusions.

106 **Material and methods**

107 **A model of optimised collections**

108 *Ex situ* conservation decisions may be driven by several criteria including location-
109 specific biodiversity conservation targets, species or breed endangerment status, and
110 economic and socio-cultural weights (i.e. which breeds are more valuable for productive
111 and other physiognomic traits). Logistical considerations are also inevitable and include
112 conservation budgets or collection and maintenance costs of gene banks (GBs). The
113 management decision may be further complicated by rationalisation options; i.e. keeping
114 fewer collections at one or several geographical locations. This objective might be
115 reasonable in the context of a collaborative research network where the free exchange
116 and use of materials are a common goal. This study the has two objectives: (i) to identify
117 least-cost logistical strategies for material collection and storage; and (ii) to estimate
118 diversity costs for different livestock species, measured as the cost per conserved breed.
119 Mathematical modelling can be used to rationalise multiple genetic collections covering
120 overlapping material at several locations.

121

122 **Model overview**

123 To represent *ex situ* diversity and logistical decisions in MIP analysis we use two
124 objective functions, namely cost minimization and diversity maximization. The latter is

125 defined as the number of breeds that are collected and stored across a set of GBs. This is
126 akin to a typical *facility location problem* (Geoffrion & Graves, 1974), which consists of
127 selecting distribution centres along with their associated customer zones. The objective is
128 to select facility sites to minimize distribution costs of demanded products. These costs
129 typically include a part that is proportional to the sum of the distances from customer
130 zones to the servicing facilities, plus costs of opening facilities at the chosen sites. The
131 facilities may or may not have limited service capacities, which in turn distinguishes the
132 problems in terms capacitated and uncapacitated variants. This problem is usually
133 modelled as NP-hard, meaning it is computationally difficult to solve, typically requiring
134 specific algorithms (Shen et al, 2011). In this application we formulate the analogous
135 problem as a simple MIP, where the facilities and customer zones are respectively
136 analogous to the gene banks and farm zones containing the breeds demanded to be
137 conserved in gene banks.

138 Figure 1 summarizes the MIP conceptualization and sets out the logistical decision
139 process for selecting breeds for *ex situ* conservation.

140 ***[Figure 1]***

141 Let BANKS 1-3 represent the GBs involved in a collective conservation program.
142 BANK 1, located in region 1 is constrained to collect breeds that are in region 1 covering
143 an average distance to farm locations, $d_1$. Alternatively, BANK 1 can collect breeds at
144 collection points, in this case $d_1=0$; i.e. breeds are brought to the bank for genetic
145 material collection depending on travel and associated logistical costs. BANK 1 can also

collect genetic materials from breeds in region 2 or region 3 by travelling $2D_{1,2} + 2d_2$ or $2D_{1,3} + 2d_3$, respectively. Collection decisions for BANK 2, BANK 3, and generalization to *n* banks is analogous to the BANK 1 case. The collection and storage of genetic materials from breeds is constrained to breed region-specific availability (i.e. whether the breed is native to the region of the specific GB), budget limitations, limited cryogenic tank capacity, maintenance and freezing costs of genetic material, and distances between gene banks and farm zones. The model currently only considers semen collection, accounting for over 90% of materials stored in the GBs covered in this study (Passemard et al., 2018)

**Model description**

The MIP model is described in terms of sets and indexes, economic parameters, decision variables, objectives functions (OFs) and constraints. Table 1 details model sets, parameters and decision variables. The model is a single-objective optimization problem, as opposed to multi-objective formulation; meaning the costs minimization of (1a) and diversity maximization (1b) are considered separately. Model equations are as follows.

***[Table 1]***

(1a) Min Costs $= \sum_{gb} C_{gb}$

(1a) represents the least collection costs objective function, given by minimizing the sum of individual gene banks costs, $C_{gb}$.

(1b) Max Diversity $= \sum_{b} \sum_{gb} SB_{b,gb}$

Equation 1b represents the diversity objective function, defined as the sum over all breeds across the gene banks; where $SB_{b,gb}$ represents the number of semen doses of breed *b* stored in gene bank *gb*.

(2) $$SB_{b,gb} = \sum_j (SP_{b,gb,j} + SF_{b,gb,j})$$

Equation 2 defines $SB_{b,gb}$ in terms of semen *b* collected by gene bank *gb* at a collection point in region *j* ($SP_{b,gb,j}$), or semen *b* collected by gene bank *gb* at farm zone *j* ($SF_{b,gb,j}$)

(3) $$SB_{b,gb} = \begin{cases} 0 \\ \mu_b \leq SB_{b,gb} \leq M_b \end{cases}$$

Equation 3 adds a further constraint to $SB_{b,gb}$ by imposing the model collects zero, or a value that is in a pre-defined interval between a minimum number of semen doses of breed *b* ($\mu_b$) and no more than $M_b$ doses.

(4) $$C_{gb} = MC_{gb} + VC_{gb}$$

Equation 4 defines the total collection and storage costs of a gene bank *gb* as the sum of maintenance costs ($MC_{gb}$) and collection costs ($CV_{gb}$).

(5) $$MC_{gb} = T * (F_{gb} + mc_{gb} \sum_b SB_{b,gb})$$

Equation 5 defines fixed maintenance costs of a gene bank *gb* as the sum of fixed annual costs ($F_{gb}$) and variable maintenance costs given by the product of storing and freezing

184 costs of gene bank *gb* ($mc_{gb}$) and the total number of semen doses in gene bank *gb* (Σ
185 $SB_{b,gb}$).

186 (6) $$VC_{gb} = CC_{gb} + TC_{gb}$$

187 Equation 6 defines variable costs as a combination of collection of gene bank *gb* ($CC_{gb}$)
188 and travel costs from gene bank *gb* ($TC_{gb}$).

189 (7) $$CC_{gb} = \sum_b (r_b cf_b \sum_j SF_{b,gb,j} + r_b cc_b \sum_j SP_{b,gb,j})$$

190 Equation 7 represents collection costs in terms of semen doses collected at farm zones,
191 first term in the sum, and at collection points, second term. The first term defines farm
192 collection costs as the total number of breeds *b* collected by gene bank *gb* at all farm
193 zones *j*; sum over *j* ($\Sigma\, SF_{b,gb,j}$), multiplied by the required number of animal donors per
194 semen dose ($r_b$) and semen collection costs per dose ($cf_b$) of breed *b*. _

195 (8) $$TC_{gb} = \frac{1}{k_{gb}} tc_{gb} \sum_b \sum_j (d_j + D_{gb,j}) SF_{b,gb,j} + \frac{1}{k_{gb}} tc_{gb} \sum_b \sum_j D_{gb,j} SP_{b,gb,j}$$

196 Equation 8 describes total travels costs of collections by gene bank *gb* ($TC_{gb}$) given by
197 the costs of collections at farm zones and at collection points, respectively the first and
198 second terms in the right hand side (RHS) of Eq. 8. The first term in the RHS defines the
199 cost of all collections of breeds *b* by gene bank *gb* across all regions *j* (double sum over *b*
200 and *j*). The multiplying parameters inside the sum, $d_j$ and $D_{gb,j}$ represent the average
201 distance from region of gene bank *j* to farm zones and the distance from gene bank *gb* to
202 gene bank *j*, respectively. The parameter $tc_{gb}$ represents the average costs per unit of

distance; $k_{gb}$ represents the average number of semen doses collected per journey by gene bank *gb*. The second term in the RHS is analogous to the first term but with $d_j = 0$.

(9) $$\sum_{gb} \sum_j e_{b,j}(SF_{b,gb,j} + SP_{b,gb,j}) = \sum_{gb} A_{b,gb}$$

Equation 9 is used to constrain the model to collect breeds that are currently available in the regional vicinity of the gene banks. The parameter $e_{b,j}$ is binary vector indicating if breed *b* is available for collection in the region of gene bank *gb* (1 if available, zero otherwise); $A_{b,gb}$ corresponds to the number of doses of breed *b* currently stored in gene bank *gb*.

(10) $$\sum_b SB_{b,gb} \leq cp_{gb}$$

Equation (10) represents gene bank capacities in number of doses of the cryogenic tanks, the total number of doses stored in gene bank *gb* ($\sum SB_{b,gb}$) cannot be greater than the capacity of gene bank *gb*, $cp_{gb}$.

(11) $$\sum_{gb} C_{gb} \leq TB$$

Equation 11 is the collection budget constraint for the gene banks, where *TB* represents the total European collective budget.

**Discontinuous variables**

Equation (3) introduces a discontinuous variable, $SB_{b,gb}$, which increases the model solving complexity and breaks its linearity. The value of $SB_{b,gb}$ must be either zero or between a particular bound. This is a necessary assumption as the number of stored doses

of semen can be zero; but if greater than zero, needs to be between an interval, for example greater than 400 doses but less than 600 doses for cattle breeds. We use a linear programming trick (Bisschop 2018) to model this discontinuous variable by introducing the indicator variable $y_b$ linked to $SB_{b,gb}$:

$$y_b = \begin{cases} 0, \text{ for } \sum_{gb} SB_{b,gb} = 0 \\ 1, \text{ for } \mu_b \leq \sum_{gb} SB_{b,gb} \leq M_b \end{cases} \quad (12)$$

The following set of constraints is used to create the desired properties in (12):

$$\sum_{gb} SB_{b,gb} \leq M_b y_b \quad (13)$$

$$\sum_{gb} SB_{b,gb} \geq \mu_b y_b \quad (14)$$

$$y_b \text{ binary} \quad (15)$$

The model was written in AIMMS algebraic language (Bisschop, 2018), comprising approximately 130,000 variables and 13,000 constraints, written in the matrix form, considering 11 selected European GBs and 489 breeds. It was solved using the CPLEX solver (IBM, 2009).

**Data**

We obtained model data from two online surveys administered to European institutions holding germplasm and genomic collections as part of the IMAGE project. The first gathered information on species germplasm and breeds. The second focused on cost data, including maintenance, costs related to semen collection and freezing, labour,

documentation, average distance between banks to farm zones, costs of skilled labour, materials and equipment and collection failure rates. The information covered six species, namely cattle, sheep, goat, horse, pig and poultry. Eleven banks returned complete information for our analysis (table 1 and table 2).

Table 3 describes the collection costs at farm zones comprising management, labour, veterinary and semen freezing costs. Table 4 presents the distances between the GBs considered in this study.

***[Table 2]***

***[Table 3]***

***[Table 4]***

## Scenarios and sensitivity analysis

The baseline scenario $S_0$ represents the current configuration of breed collections distributed across the 11 GBs. The analysis assumes that breeds currently available in each bank are native to the respective region; i.e., if a given cattle breed *b1* is stored in BANK 1 in $S_0$, it can only be collected from region 1. The optimized scenario $S_{UC}$ represents the minimum cost breed-gene bank allocation allowing cross-region collection, and assumes hypothetically an unlimited capacity of cryogenic tanks. $S_{C50}$ is analogous to $S_{UC}$ but assumes that current bank breeds occupy 50% of tank capacities. $S_1$, $S_2$,..., $S_{11}$ assume unlimited capacities and impose one bank to store all breeds. That is, $S_1$ means all breeds across the 11 regions are collected by BANK 1 only, and analogously for $S_2$ to $S_{11}$.

While the previous scenarios are explored by minimizing collection costs, a sensitivity analysis employs a diversity function (Eq. 1b) to estimate the cost per conserved breed for cattle, sheep, goat, poultry and pigs, depending on available budget. We define a diversity value ($D_i$), measured by the number of selected breeds, as a function of $n$ available budgets ($B_i$) as follows:

$$B_i = B_0 + i\frac{(B_{Max} - B_0)}{n} \quad (16)$$

Where $B_0$ represents the initial budget; $i$ represents the budget scenario and varies from 1 to $n$=100 (an arbitrary number of scenarios); and $B_{Max}$ is the maximum available budget. For all $B_i$'s, the value of diversity is calculated as $D_i$ = f($B_i$). Where f($B_i$) represents the

optimal solution of the MIP model when maximizing the diversity function. $D_i$ and $B_i$ are plotted as y-and x-axis to define efficient cost curves for each of the livestock species.

## Results

The estimated total cost of the current breed allocations for the 11 European GBs is 23.2 M EUR, including regional collection costs only because the $S_0$ scenario assumes no cross-regional collection strategies. Of the 489 breeds, 55% of semen doses are cattle, 25% are sheep, 9% pig, 4% poultry, followed by goat and horse, together representing 3% of the total collected doses. Figure 2 shows how breeds, according to species, are distributed across the GBs. Most cattle breeds are distributed between B1, B6, B7 and B9 to B11. While 78% of sheep breeds are in B1, B6 and B9. Pig breeds are mostly in B7 (82%). Analysis of $S_0$ shows overlapping collection of cattle, goat, sheep and pigs, and most significantly for cattle; e.g., the cattle breed Blonde D'Aquitaine is currently stored in five different banks, varying form 50 semen doses (B11) to 9,670 (B1) (see appendix table A1 for details).

***[Figure 2]***

Figure 3 shows the alternative least cost collection strategy ($S_{UC}$). If the breeds in figure 2 were collected at least cost, the required budget would fall by 23%, or around 5.4 M EUR. Figure 3 shows $S_{UC}$ reduces collection costs by transferring cattle breeds from GBs 6, 7,9 and 11 to GBs 1, 2 and 10, relative to $S_0$, sheep breeds from GB 3 to GB 11, horse

293 breeds from B7 to B1, while poultry and pigs are kept approximately the same as $S_0$. In
294 fact, Table 2 shows those cryogenic banks have relatively low fixed and collection costs,
295 and are geographically closer to other collection regions (as shown in Table 4).
296
297 ***[Figure 3]***
298
299 Figure 4 shows that assuming the GBs are currently operating at 50% of their capacities
300 ($S_{C50}$) reduces the required budget by 19% relative to $S_0$, or by 4.3 M EUR. Figure 4
301 shows that the strategy for minimising costs under limited capacity, in relation to $S_{UC}$, is
302 transferring cattle breed doses from B2 and B10 to B6 and B11 due to the first two being
303 above the gene banks capacity. This is because under $S_{UC}$ only BANK 2 exceeds the
304 assumed tank capacity, thus moving to the closest GB, BANK 11, which is only 39.7 km
305 away (Table 4) minimises collection costs.
306
307 ***[Figure 4]***
308
309 The alternative single gene bank scenarios $S_1$ to $S_{11}$ are presented in figure 5, showing
310 costs varying by +100% to 285% relative to $S_0$. The difference is explained by an
311 inefficient collection strategy that ignores the relative breed costs across different GBs
312 and the variation in travel costs, which are in turn related to the number of doses
313 collected by a bank per visit to a region and farm zone.
314

*[Figure 5]*

Figure 6 shows diversity cost curves for cattle, sheep, goat, poultry, pigs and horse. Increasing marginal cost per stored breed reflects the fact that breed collection takes place in the same region as the banks (i.e. native breeds). As the available budget (or cost) increases (x-axis), more genetic material from more breeds can be collected, although at higher cost and requiring more cross-regional collection due to cryogenic tank capacity limitations (upper cost) up to the point where all available breeds within the GBs are collected, as the model assumes unlimited tank capacity for constructing the diversity cost curves. Collection and storage cost per breed varies between 55 and 2531 EUR, depending on the number of breeds and species that are already collected.

*[Figure 6]*

## Discussion

Model results suggest a potential for cost saving across European cryogenic banks by strategic collection and conservation planning. The results indicate overlapping collections across the GBs. However, from a genetic viewpoint the same breed in different countries may harbour different genetic diversities. Despite we do not consider transboundary breeds those won't change the results as they represent a small fraction of the total breeds considered in this study (21 out of 489).

As well as being risky, collecting all materials in one of the existing cryogenic banks considered in this study is a relatively expensive option compared with multiple banks. This is because there is no GB with storage and collection costs cheaper for all species, and storing in a single GB increases cross-country transportation costs because some breeds are unavailable for collection in a single region.

Our results could be refined by improved cost estimates for collections and the addition of more breeds and country collections not taken into account in the analysis, e.g. from FAO DAD-IS database (FAO-DADIS, 2018). However, adding these will also require collection of further regional cost data, which are usually incomplete. For example the cost surveys used in this study (Passemard et al., 2018) revealed inconsistent approaches to the recording of collection and storage costs, and different ways of recording units of collected materials.

A further apparent limitation is the representativeness of the cost survey, or more specifically varying banking technologies. A bank with lower storage costs may imply poorer conservation quality and higher failure risks, which we would ideally include as a specific variable in the optimisation problem.

The analysis further suggests that the diversity-cost curve varies for breeds of each species. Since the surveys did not seek to understand breed-specific weighting criteria, the diversity-cost curves do not discriminate in terms of important breed-specific attributes, e.g related to greater or lower expected economic returns, breed endangerment, susceptibility to climate change or cultural attributes.

Like relative failure risks, these attributes could be included in the MIP as stochastic parameters, for example by forcing the model to select breeds with probability of extinction greater than a threshold value, variance of expected return, or probability of successful restoration. Assignment of breed-specific and technology attribute weights to the diversity function will change the shape of the curves.

Many breed attributes are likely to be prominent in any survey of public preferences for conservation spending, which we suspect would mirror priorities for public good provision related to *in situ* conservation decisions. However, this raises a further question about the demand for *ex situ* collections, and whether they should serve public or private good objectives; the latter focussed largely on animal productivity traits. This largely depends on ownership and how they are financed. The survey by Passmard et al (2018) suggested that most of the respondent collections were under the auspices of public institutions. However, the same survey did not seek views on how this translates into objectives for attribute selection. Furthermore, the survey most likely overlooks other genetic material held in private (i.e. industrial) collections. Accordingly, without clarity on private sector preferences we can only optimise over a known proportion of the stored resources.

## Conclusion

Rapid progress in the development of next generation gene sequencing and bioinformatic tools have revolutionised animal breeding, but potentially distracted from a basic problem of what genetic and reproductive materials to collect and store, and how stored

information is consistently recorded. Breed and gene bank selection clearly involves numerous biotechnological, institutional and economic challenges that can be informed by mathematical modelling of cost-effective breed conservation.

For given objectives and constraints our model provides some indication of potential rationalisation options and demonstrates the increasing marginal costs of conservation effort. The exercise begs important questions about the specific optimisation objectives, which in turn require more institutional coordination to define the mix of private and public good objectives and hence potential cost and benefit sharing. This implies clearer articulation of *in situ* risks including endangerment due to climate change and other pressures, expected economic returns and other attributes that determine stakeholders' conservation preferences. There is also a need to improve understanding of the efficacy of technologies developed for *ex situ* curation and associated risks of successful use in future agricultural scenarios.

**Acknowledgements**

We acknowledge the participating institutions that provided cost data for this work. We acknowledge funding from the IMAGE (Innovative Management of Animal Genetic Resources) project funded under the European Union's Horizon 2020 research and innovation program under grant agreement No 677353. We also acknowledge funding from The Scottish Government Strategic Research Programme 2016-2021, work package 2.3 Agricultural Systems and Land Management.

## Conflict of Interest Statement

The authors declare no conflict of interest.

## References


Bisschop, J. (2018). *AIMMS Optimization Modeling*. *AIMMS Optimization Modeling*. https://doi.org/10.1002/9781118267752.ch5

Blackburn, H. D. (2009). Genebank development for the conservation of livestock genetic resources in the United States of America. *Livestock Science*, *120*(3), 196–203. https://doi.org/10.1016/j.livsci.2008.07.004

FAO-DADIS. (2018). Domestic animal diversity information system. Retrieved November 1, 2018, from http://www.fao.org/dad-is/browse-by-country-and-species/en/

FAO. (2009). *International treaty on plant genetic resources for food and agriculture*. Rome. Retrieved from http://www.fao.org/3/a-i0510e.pdf

FAO. (2015). *The Second Report on the State of the World's Animal Genetic Resources for Food and Agriculture*. (S. B. D. & P. D., Ed.). Rome (Italy): FAO Commission on Genetic Resources for Food and Agriculture Assessments. Retrieved from http://www.fao.org/3/a-i4787e/index.html

Geoffrion, A. M., & Graves, G. W. (1974). Multicommodity Distribution System Design by Benders Decomposition: Part II. *Management Science*. https://doi.org/10.1287/mnsc.20.5.822

Groeneveld, L. F., Gregusson, S., Guldbrandtsen, B., Hiemstra, S. J., Hveem, K., Kantanen, J., … Berg, P. (2016). Domesticated Animal Biobanking: Land of Opportunity. *PLoS Biology*, *14*(7). https://doi.org/10.1371/journal.pbio.1002523

Hiemstra, S. J., Martyniuk, E., Duchev, Z., & Begemann, F. (2014). European Gene Bank Network for Animal Genetic Resources (EUGENA). *Procceedings of the 10th WCGALP*, (Mackay), 10. Retrieved from https://www.asas.org/docs/default-source/wcgalp-posters/437_paper_8691_manuscript_289_0.pdf?sfvrsn=2

IBM. (2009). V12. 1: User's Manual for CPLEX. *International Business Machines Corporation*, *46*(53), 157.

Kukkala, A. S., & Moilanen, A. (2013). Core concepts of spatial prioritisation in systematic conservation planning. *Biological Reviews*, *88*(2), 443–464. https://doi.org/10.1111/brv.12008

Lee, J., & Letchford, A. N. (2007). Mixed integer programming. *Discrete Optimization*, *4*(1), 1–13. https://doi.org/10.1016/j.disopt.2006.10.005

McGowan, P. J. K., Traylor-Holzer, K., & Leus, K. (2017). IUCN Guidelines for

Determining When and How Ex Situ Management Should Be Used in Species Conservation. *Conservation Letters*, *10*(3), 361–366. https://doi.org/10.1111/conl.12285

Önal, H. (2003). Preservation of Species and Genetic Diversity. *American Journal of Agricultural Economics*, *85*(2), 437–447. https://doi.org/10.1111/1467-8276.00132

Passemard, A., Joly, L., Duclos, D. & Danchin-Burge, C. (2018). *Inventory and mapping of European animal genetic collections, IDELE, IMAGE (Innovative Management of Animal Genetic Resources) project report.*

Pizzi, F., Turri, F., Gliozzi, T. M., & Gandini, G. (2016). Implementation and cost analysis of a regional farm animal cryobank: An Italian case study. *Italian Journal of Animal Science*, *15*(2), 207–210. https://doi.org/10.1080/1828051X.2016.1153406

Reist-Marti, S. B., Abdulai, A., & Simianer, H. (2006). Optimum allocation of conservation funds and choice of conservation programs for a set of African cattle breeds. *Genetics Selection Evolution*, *38*(1), 99. https://doi.org/10.1186/1297-9686-38-1-99

Shen, Z. J. M., Zhan, R. L., & Zhang, J. (2011). The reliable facility location problem: Formulations, heuristics, and approximation algorithms. *INFORMS Journal on Computing*, *23*(3), 470–482. https://doi.org/10.1287/ijoc.1100.0414

463 **Tables**

464

465 **Table 1. Model description in terms of sets, economic parameters and decision**

466 **variables.**

| Model inputs | Description | |
|---|---|---|
| ***Sets*** | | |
| B | Breed | |
| Gb | gene bank | |
| | | |
| ***Parameters*** | Description | Value |
| T | storage time | 1 year |
| $F_{gb}$ | gene bank *gb* maintenance cost | |
| $mc_{gb}$ | gene bank *gb* liquid N and storing costs | Table 2 |
| $F_{gb}$ | fixed maintenance costs of gene bank *gb* | |
| TB | total conservation budget | 600 to 118800 EUR |
| $cf_b$ | semen collection cost of gene bank g*b* at farm | Table 3 |
| $cc_b$ | semen collection cost of gene bank *gb* at collection point | Table 3 |
| $tc_{gb}$ | travel costs of gene bank *gb*, proportional to distance | 2.5 €.km$^{-1}$ |
| $d_{gb}$ | distance between gene bank *gb* and farm zone containing breed *b* samples | Table 2 |
| $D_{gb,j}$ | distance between gene bank *gb* and gene bank *j*, where *j* is an auxiliary index for gene banks | Table 4 |
| $A_{b,gb}$ | number of doses of breed *b* currently stored in gene bank *gb* | Passemard et al. (2018) |
| $r_b$ | number of required samples of breed *b* for collection of semen doses | 25 animals |
| $cp_{gb}$ | capacity of cryotanks in gene bank *gb* | Table 2 |
| $\mu_b$ | minimum collection in number of doses of semen of breed *b* | 0.8 of $M_b$ |
| $M_b$ | Maximum collection in number of doses of semen of breed *b* | cattle (585), sheep(198), goat(105), horse(58), pig(150), poultry(46) |

| | | |
|---|---|---|
| $e_{b,gb}$ | binary parameter indicating if breed *b* is available for collection in region of gene bank *gb* | Passemard et al. (2018) |
| $k_{gb}$ | number of doses that gene bank *gb* collects per travel | 400 doses |
| ***Decision variables*** | | |
| $SB_{b,gb}$ | amount of doses of semen breed *b* stored in gene bank *gb* | |
| $SF_{b,gb,j}$ | amount of semen of breed *b* collected by gene bank *gb* at a farm zone in region *j* | |
| $SP_{b,gb,j}$ | amount of semen of breed *b* collected by gene bank *gb* at a collection point in region *j* | |
| $FC_{gb}$ | total fixed costs of gene bank *gb* | |
| $VC_{gb}$ | total variable costs (collection) of gene bank *gb* | |
| $TC_{gb}$ | variable travels costs (collection) of gene bank *gb* | |
| $CC_{gb}$ | variable collections costs (collection) of gene bank *gb* | |
| $C_{gb}$ | total costs of gene bank *gb* | |

[a] Each GB is associated with its regional vicinity, thus the set of GBs is equivalent to the set of regions. The index *j* is used as an auxiliary to *gb*.

**Table 2. Input data used in the model including the cost parameters, tank capacities and distances.**

| Gene banks | Location | Maintenance cost,$F_{gb}$ (1000 €.yr$^{-1}$) | Storage cost,$mc_{gb}$ (€.dose$^{-1}$) | Tanks capacity[1], $C_{gb}$ (doses) | Doses currently stored, $\Sigma_b A_{b,gb}$ (doses) | Distance to farm zones[2], $d_{gb}$ (km) |
|---|---|---|---|---|---|---|
| B1 | Paris, France | 119 | 0.12 | 607776 | 303888 | 200 |
| B2 | Madri, Spain | 41 | 0.56 | 55160 | 27580 | 300 |
| B3 | Valdepenas, Spain | 45 | 0.30 | 88120 | 44060 | 200 |
| B4 | Bellaterra, Spain | 115 | 1.89 | 10946 | 5473 | 200 |
| B5 | Godollo, Hungary | 38 | 3.88 | 4124 | 2062 | 200 |
| B6 | Thalheim, Gemany | 338 | 0.23 | 435174 | 217587 | 100 |
| B7 | Wageningen, Netherlands | 190 | 0.05 | 882470 | 441235 | 100 |
| B8 | P. de Mallorca, Spain | 36 | 0.97 | 30018 | 15009 | 100 |
| B9 | Kenilworth, UK | 115 | 0.15 | 551944 | 275972 | 500 |
| B10 | Kiev, Ukraine | 115 | 0.07 | 292602 | 146301 | 100 |
| B11 | Colmenar V., Spain | 115 | 0.15 | 335350 | 167675 | 200 |

[a] Fixed maintenance costs are composed of labour, property rent and depreciation of tanks and equipment.

[b] Liquid N and other storing costs.

[c] Assumed as 50% of current usage.

[d] Round trip distance.

**Table 3. Collection costs at farm zones and collection points, comprising management, labour, veterinary and semen freezing costs.**

| | Collection costs (EUR.animal$^{-1}$) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | At farm zones, $cf_b$ | | | | | | At collection centre, $cc_b$ | | | | | |
| | Cattle | Sheep | Goat | Horse | Pig | Poultry | Cattle | Sheep | Goat | Horse | Pig | Poultry |
| B1 | 100 | 100 | 100 | 150 | 340 | 15 | 182 | 30 | 30 | 50 | 30 | 15 |
| B2 | 50 | 100 | 100 | 150 | 340 | 30 | 50 | 30 | 30 | 50 | 30 | 30 |
| B3 | 100 | 100 | 100 | 150 | 340 | 15 | 182 | 30 | 30 | 50 | 30 | 15 |
| B4 | 100 | 100 | 100 | 150 | 340 | 15 | 182 | 30 | 30 | 50 | 30 | 15 |
| B5 | 100 | 100 | 100 | 150 | 340 | 1 | 182 | 30 | 30 | 50 | 30 | 1 |
| B6 | 100 | 100 | 100 | 150 | 340 | 15 | 182 | 30 | 30 | 50 | 30 | 15 |
| B7 | 300 | 150 | 150 | 300 | 150 | 15 | 300 | 150 | 150 | 300 | 150 | 15 |
| B8 | 60 | 30 | 40 | 60 | 40 | 15 | 60 | 30 | 30 | 60 | 40 | 15 |
| B9 | 500 | 100 | 100 | 150 | 340 | 15 | 500 | 30 | 30 | 50 | 30 | 15 |
| B10 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 |
| B11 | 100 | 100 | 100 | 150 | 340 | 15 | 182 | 30 | 30 | 50 | 30 | 15 |

**Table 4: Relative distances between the gene banks (in km).**

| | Distance between gene banks, $D_{gb,j}$ (km) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | B1 | B2 | B3 | B4 | B5 | B6 | B7 | B8 | B9 | B10 | B11 |
| B1 | 0 | 1274 | 1478 | 1028 | 1538 | 1045 | 495 | 1290 | 640 | 2402 | 1238 |
| B2 | 1274 | 0 | 214 | 611 | 2572 | 2213 | 1761 | 740 | 1900 | 3687 | 39.7 |
| B3 | 1478 | 214 | 0 | 682 | 2644 | 2285 | 1963 | 679 | 2103 | 3789 | 250 |
| B4 | 1028 | 611 | 682 | 0 | 1962 | 1602 | 1498 | 290 | 1658 | 3107 | 631 |
| B5 | 1538 | 2572 | 2644 | 1962 | 0 | 501 | 1374 | 2247 | 1950 | 1094 | 2593 |
| B6 | 1045 | 2213 | 2285 | 1602 | 501 | 0 | 1027 | 2034 | 1603 | 1362 | 2379 |
| B7 | 495 | 1761 | 1963 | 1498 | 1374 | 1027 | 0 | 1780 | 697 | 1912 | 1731 |
| B8 | 1290 | 740 | 679 | 290 | 2247 | 2034 | 1780 | 0 | 1920 | 3369 | 728 |
| B9 | 640 | 1900 | 2103 | 1658 | 1950 | 1603 | 697 | 1920 | 0 | 2578 | 1870 |
| B10 | 2402 | 3687 | 3789 | 3107 | 1094 | 1362 | 1912 | 3369 | 2578 | 0 | 3657 |
| B11 | 1238 | 39.7 | 250 | 631 | 2593 | 2379 | 1731 | 728 | 1870 | 3657 | 0 |

**Figures**

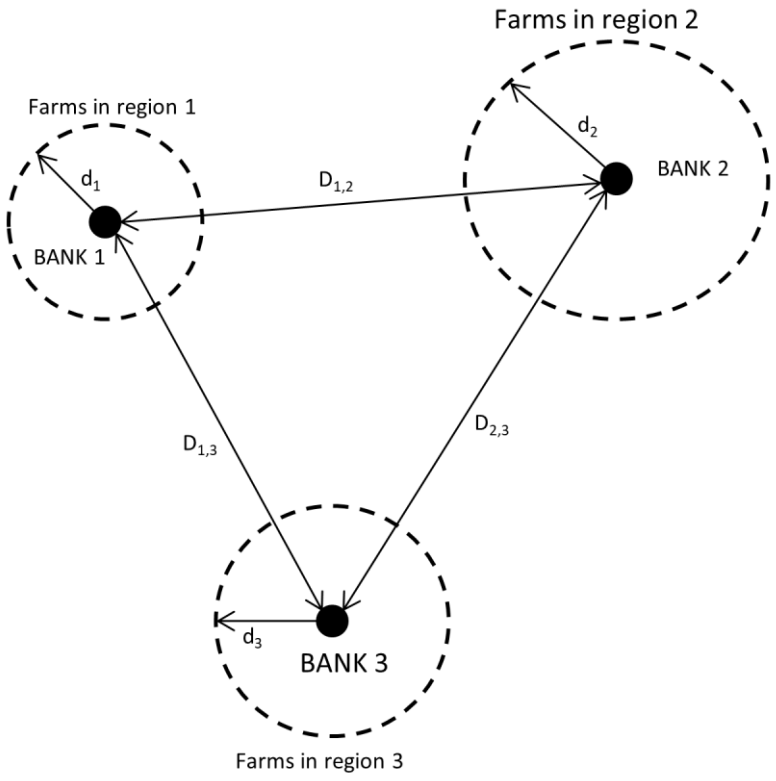



**Figure 1: Conceptualization of gene bank optimization considering three illustrative gene banks (BANK 1-3), distance between the banks (e.g., $D_{1,2}$) and distance from gene banks to farm zones, e.g., $d_1$.**

**Figure 2: Baseline ($S_0$) gene bank allocation (total number of doses within a species) across the 11 European banks for livestock breeds.**

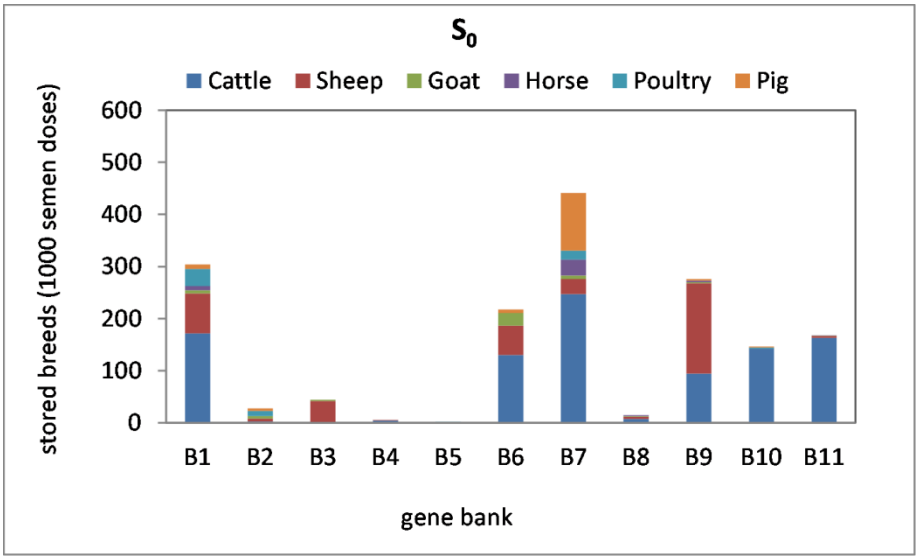


**Figure 3: Least cost gene bank allocation (in total number of doses within a species) across the selected 11 European banks for livestock breeds under unlimited capacity of cryogenic tanks, scenario $S_{UC}$.**

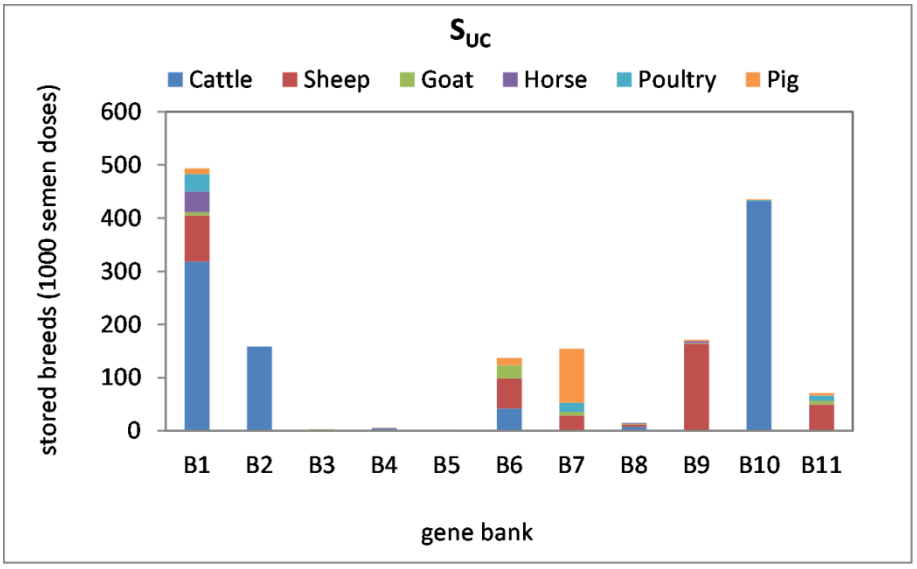


**Figure 4: Least cost gene bank allocation (in total number of doses within a species) across the selected 11 European banks for livestock breeds under limited capacity of cryogenic tanks, scenario $S_{C50}$.**

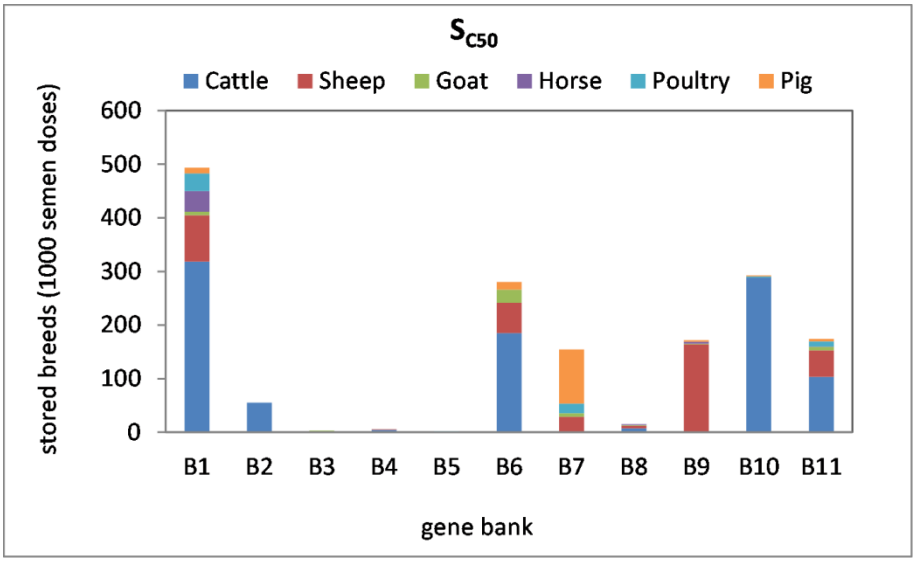


**Figure 5: Gene bank allocation scenarios and associated costs. S0, SC50 and SUC are the current configuration of breed collections, constraint and unconstraint minimum cost collection scenarios, respectively. Scenarios S1 to S11 represent single gene bank allocations; S1 means all breeds are collected by BANK 1 only, and analogously from S2 to S11.**

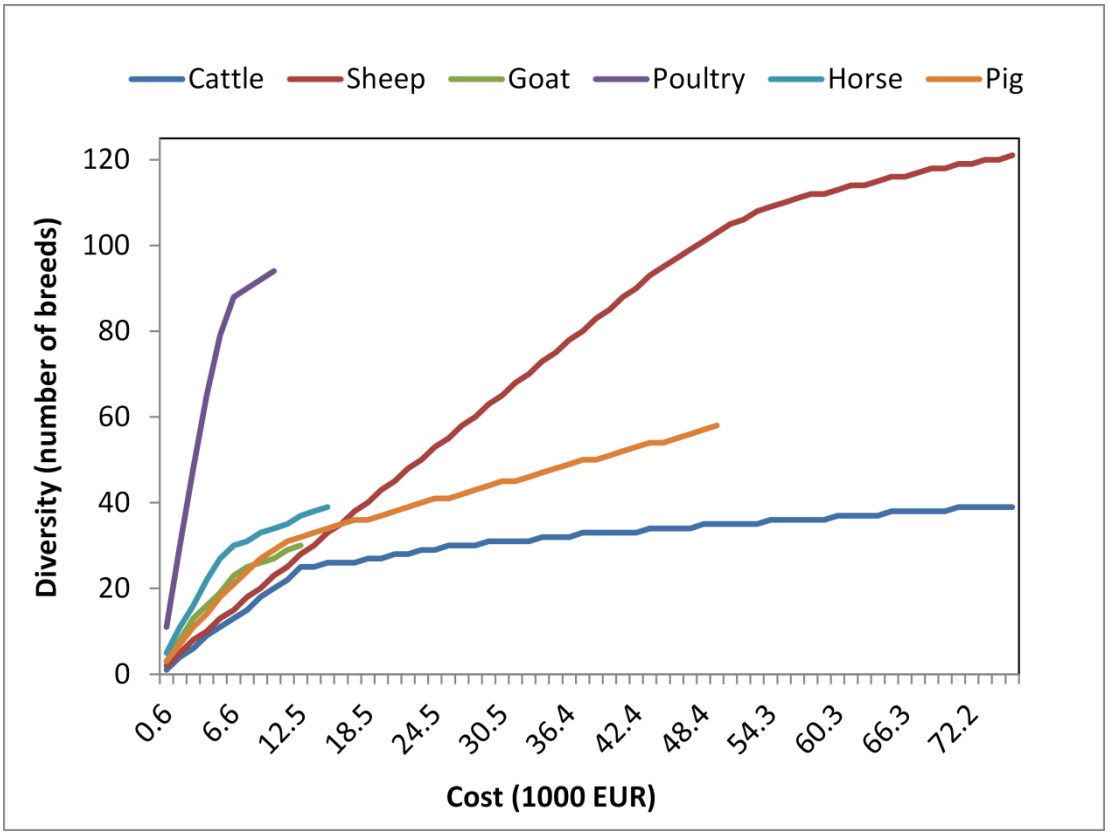
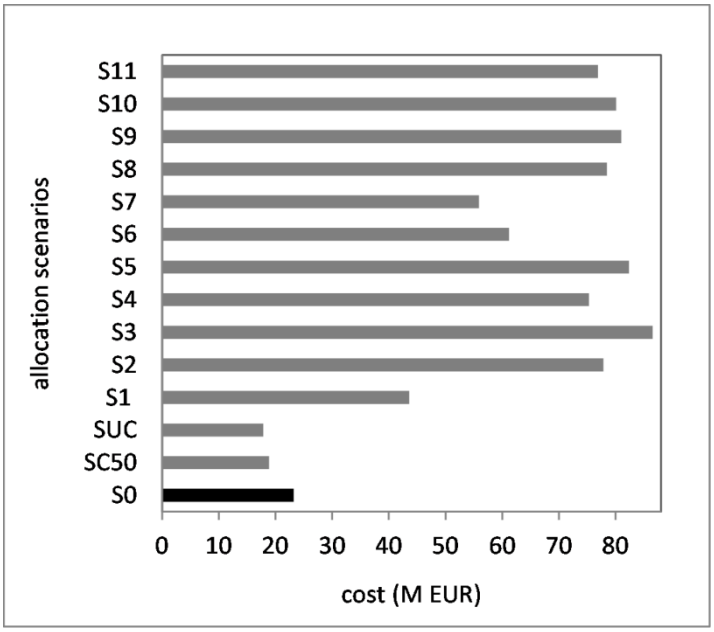


**Figure 6: Sensitivity analysis of diversity as a function of collective EU budget for livestock breeds.**



## Appendix

### Table A1: Overlapping breeds and number of semen doses in the GBs.

| | Number of semen doses | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Breed** | B1 | B2 | B3 | B4 | B5 | B6 | B7 | B8 | Total |
| Cattle - Belgian Blue | | | | 1150 | 375 | | | | 1525 |
| Cattle - Blonde D´aquitaine | 9670 | | | 350 | 75 | | 770 | 50 | 10915 |
| Cattle - Brown Swiss | | | | 1534 4 | 87 | | | | 15431 |
| Cattle - Charolaise | 1160 0 | | | 672 | | | 1649 | 439 6 | 18317 |
| Cattle - Galloway | | | | 100 | | 711 | | | 811 |
| Cattle - Hereford | | | | | | 486 | 2000 | | 2486 |
| Cattle - Holstein | | | | | 14294 4 | | 3604 0 | | 65547 |
| Cattle - Jersey | | | | | 100 | | 1050 | | 1150 |
| Cattle - Limousine | 7000 | | | 1650 | | | 3539 | 244 7 | 14636 |
| Cattle - Montbeliard | 2110 0 | | | 92 | 75 | | 218 | | 21485 |
| Cattle - Piedmont | | | | 100 | 25 | | 3000 | | 3125 |
| Cattle - Simmental | | | | 8620 0 | 25 | | 1691 4 | | 10313 9 |
| Goat - Murciano Granadina | | | 1337 | | | | | 43 | 1380 |
| Goat - Saanen | 923 | | | | 75 | | | | 998 |
| Pig - Duroc | 287 | | | | 2378 | | | | 2665 |
| Pig - Landrace | 298 | | | 200 | | | | | 498 |
| Pig - Large White | | | | 134 | | 250 | | | 384 |
| Pig - Pietrain | | | | 602 | 7033 | | | | 7635 |
| Sheep - Manchega | | 72 5 | 3979 4 | | | | | 304 3 | 43562 |
| Sheep - Romaney | 2534 | | | | | 240 2 | | | 4936 |
| Sheep - Suffolk | 5509 | | | | | 743 4 | | | 12943 |